CHƯƠNG V
THỐNG KÊ

I. Một số khái niệm

- Một tập con hữu hạn các đơn vị điều tra đgl một **mẫu**.
- Số phần tử của một mẫu đgl **kích thước mẫu**.
- Các giá trị của dấu hiệu thu được trên mẫu đgl một **mẫu số liệu**.

II. Trình bày một mẫu số liệu

- **Tần số** của một giá trị là số lần xuất hiện của mỗi giá trị trong mẫu số liệu.
- **Tần suất** f_i của giá trị x_i là tỉ số giữa tần số n_i và kích thước mẫu *N*:

$$f_i = \frac{n_i}{N}$$ *(thường viết tần suất dưới dạng %)*

- **Bảng phân bố tần số – tần suất**

Giá trị	Tần số	Tần suất (%)
x_1	n_1	f_1
x_2	n_2	f_2
...	...	...
x_k	n_k	f_k
	N	100 (%)

- **Bảng phân bố tần số – tần suất ghép lớp**

Lớp	Tần số	Tần suất (%)
$[x_1; x_2)$	n_1	f_1
$[x_2; x_3)$	n_2	f_2
...	...	...
$[x_k; x_{k+1})$	n_k	f_k
	N	100 (%)

III. Biểu đồ

- **Biểu đồ hình cột**
- **Biểu đồ hình quạt**
- **Đường gấp khúc**

IV. Các số đặc trưng của mẫu số liệu

1. Số trung bình

- Với mẫu số liệu kích thước N là $\{x_1, x_2, ..., x_N\}$:

$$\overline{x} = \frac{x_1 + x_2 + ... + x_N}{N}$$

- Với mẫu số liệu được cho bởi bảng phân bố tần số:

$$\overline{x} = \frac{n_1x_1 + n_2x_2 + ... + n_kx_k}{N}$$

- Với mẫu số liệu được cho bởi bảng phân bố tần số ghép lớp:

$$\overline{x} = \frac{n_1c_1 + n_2c_2 + ... + n_kc_k}{N}$$ *(c_i là giá trị đại diện của lớp thứ i)*

2. Số trung vị

Giả sử ta có một mẫu gồm N số liệu được sắp xếp theo thứ tự không giảm (hoặc không tăng). Khi đó **số trung vị M_e** là:

– Số đứng giữa nếu N lẻ;
– Trung bình cộng của hai số đứng giữa nếu N chẵn.

3. Mốt

Mốt của một bảng phân bố tần số là giá trị có tần số lớn nhất và được kí hiệu là M_O.

Chú ý: *– Số trung bình của mẫu số liệu được dùng làm đại diện cho các số liệu của mẫu.*
– Nếu các số liệu trong mẫu có sự chênh lệch quá lớn thì dùng số trung vị làm đại diện cho các số liệu của mẫu.
– Nếu quan tâm đến giá trị có tần số lớn nhất thì dùng mốt làm đại diện. Một mẫu số liệu có thể có nhiều mốt.

4. Phương sai và độ lệch chuẩn

Để đo mức độ chênh lệch (**độ phân tán**) giữa các giá trị của mẫu số liệu so với số

trung bình ta dùng **phương sai** s^2 và **độ lệch chuẩn** $s=\sqrt{s^2}$.

• Với mẫu số liệu kích thước N là $\{x_1, x_2, ..., x_N\}$:

$$s^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i-\overline{x})^2 = \frac{1}{N}\sum_{i=1}^{N}x_i^2 - \frac{1}{N^2}\left(\sum_{i=1}^{N}x_i\right)^2 = \overline{x^2}-(\overline{x})^2$$

• Với mẫu số liệu được cho bởi bảng phân bố tần số, tần suất:

$$s^2 = \frac{1}{N}\sum_{i=1}^{k}n_i(x_i-\overline{x})^2 = \frac{1}{N}\sum_{i=1}^{k}n_i x_i^2 - \frac{1}{N^2}\left(\sum_{i=1}^{k}n_i x_i\right)^2$$

$$= \sum_{i=1}^{k}f_i(x_i-\overline{x})^2 = \sum_{i=1}^{k}f_i x_i^2 - \left(\sum_{i=1}^{k}f_i x_i\right)^2$$

• Với mẫu số liệu được cho bởi bảng phân bố tần số, tần suất ghép lớp:

$$s^2 = \frac{1}{N}\sum_{i=1}^{k}n_i(c_i-\overline{x})^2 = \frac{1}{N}\sum_{i=1}^{k}n_i c_i^2 - \frac{1}{N^2}\left(\sum_{i=1}^{k}n_i c_i\right)^2$$

$$= \sum_{i=1}^{k}f_i(c_i-\overline{x})^2 = \sum_{i=1}^{k}f_i c_i^2 - \left(\sum_{i=1}^{k}f_i c_i\right)^2$$

(c_i, n_i, f_i là giá trị đại diện, tần số, tần suất của lớp thứ I;
N là số các số liệu thống kê $N = n_1+n_2+...+n_k$*)*

Chú ý: *Phương sai và độ lệch chuẩn càng lớn thì độ phân tán (so với số trung bình) của các số liệu thống kê càng lớn.*

Bài 1. Trong các mẫu số liệu dưới đây:

i) Cho biết dấu hiệu và đơn vị điều tra là gì? Kích thước mẫu là bao nhiêu?
ii) Lập bảng phân bố tần số, tần suất. Nhận xét.
iii) Vẽ biểu đồ tần số, tần suất.
iv) Tính số trung bình, số trung vị, mốt.
v) Tính phương sai và độ lệch chuẩn. Nhận xét.

1) Tuổi thọ của 30 bóng đèn được thắp thử (*đơn vị: giờ*)

1180	1150	1190	1170	1180	1170	1160	1170	1160	1150
1190	1180	1170	1170	1170	1190	1170	1170	1170	1180
1170	1160	1160	1160	1170	1160	1180	1180	1150	1170

2) Năng suất lúa hè thu (*tạ/ha*) năm 1998 của 31 tỉnh

30	30	25	25	35	45	40	40	35	45	
25	45	30	30	30	40	30	25	45	45	
35	35	30	40	40	40	35	35	35	35	35

3) Số con của 40 gia đình ở huyện A.

2	4	3	2	0	2	2	3	4	5
2	2	5	2	1	2	2	2	3	2
5	2	7	3	4	2	2	2	3	2
3	5	2	1	2	4	4	3	4	3

4) Điện năng tiêu thụ trong một tháng (*kW/h*) của 30 gia đình ở một khu phố A.

165	85	65	65	70	50	45	100	45	100
100	100	100	90	53	70	141	42	50	150
40	70	84	59	75	57	133	45	65	75

5) Số học sinh giỏi của 30 lớp ở một trường THPT.

0	2	1	0	0	3	0	0	1	1	0	1	6	6	0
1	5	2	4	5	1	0	1	2	4	0	3	3	1	0

6) Nhiệt độ của 24 tỉnh, thành phố ở Việt Nam vào một ngày của tháng 7 (*đơn vị: độ*)

36	30	31	32	31	40	37	29	41	37	35	34
34	35	32	33	35	33	33	31	34	34	32	35

6) Tốc độ (*km/h*) của 30 chiếc xe môtô ghi ở một trạm kiểm soát giao thông.

40	58	60	75	45	70	60	49	60	75
52	41	70	65	60	42	80	65	58	55
65	75	40	55	68	70	52	55	60	70

7) Kết quả điểm thi môn Văn của hai lớp 10A, 10B ở một trường THPT.

Lớp 10A	Điểm thi	5	6	7	8	9	10	Cộng
	Tần số	1	9	12	14	1	3	40

Lớp 10B	Điểm thi	6	7	8	9	Cộng
	Tần số	8	18	10	4	40

8) Tiền lương hàng tháng của 30 công nhân ở một xưởng may.

Tiền lương	300	500	700	800	900	1000	Cộng
Tần số	3	5	6	5	6	5	30

9) Một nhà nghiên cứu ghi lại tuổi của 30 bệnh nhân mắc bệnh đau mắt hột.

21	17	22	18	20	17	15	13	15	20	15	12	18	17	25
17	21	15	12	18	16	23	14	18	19	13	16	19	18	17

10) Năng suất lúa (*đơn vị: tạ/ha*) của 120 thửa ruộng ở một cánh đồng.

Năng suất	30	32	34	36	38	40	42	44
Tần số	10	20	30	15	10	10	5	20

Bài 2. Trong các mẫu số liệu dưới đây:

i) Cho biết dấu hiệu và đơn vị điều tra là gì? Kích thước mẫu là bao nhiêu?
ii) Lập bảng phân bố tần số, tần suất ghép lớp. Nhận xét.
iii) Vẽ biểu đồ tần số, tần suất.
iv) Tính số trung bình, số trung vị, mốt.
v) Tính phương sai và độ lệch chuẩn. Nhận xét.

1) Khối lượng của 30 củ khoai tây thu hoạch được ở nông trường T (*đơn vị: g*).

90	73	88	99	100	102	101	96	79	93
81	94	96	93	95	82	90	106	103	116
109	108	112	87	74	91	84	97	85	92

Với các lớp: [70; 80), [80; 90), [90; 100), [100; 110), [110; 120].

2) Chiều cao của 35 cây bạch đàn (*đơn vị: m*).

6,6	7,5	8,2	8,2	7,8	7,9	9,0	8,9	8,2	7,2	7,5	8,3
7,4	8,7	7,7	7,0	9,4	8,7	8,0	7,7	7,8	8,3	8,6	8,1
8,1	9,5	6,9	8,0	7,6	7,9	7,3	8,5	8,4	8,0	8,8	

Với các lớp: [6,5; 7,0), [7,0; 7,5), [7,5; 8,0), [8,0; 8,5), [8,5; 9,0), [9,0; 9,5].

3) Số phiếu dự đoán đúng của 25 trận bóng đá học sinh.

54	75	121	142	154	159	171	189	203	211	225	247	251
259	264	278	290	305	315	322	355	367	388	450	490	

Với các lớp: [50; 124], [125; 199], … (độ dài mỗi đoạn là 74).

4) Doanh thu của 50 cửa hàng của một công ti trong một tháng (*đơn vị: triệu đồng*).

102	121	129	114	95	88	109	147	118	148	128	71	93
67	62	57	103	135	97	166	83	114	66	156	88	64
49	101	79	120	75	113	155	48	104	112	79	87	88
141	55	123	152	60	83	144	84	95	90	27		

Với các lớp: [26,5; 48,5), [48,5; 70,5), … (độ dài mỗi khoảng là 22).

5) Điểm thi môn Toán của 60 học sinh lớp 10.

1	5	4	8	2	9	4	5	3	2	7	2	7	10	0
2	6	3	7	5	9	10	10	7	9	0	5	3	8	2
4	1	3	6	0	10	3	3	0	8	6	4	1	6	8
2	5	2	1	5	1	8	5	7	2	4	6	3	4	2

Với các lớp: [0;2), [2; 4), …, [8;10].

6) Số điện tiêu thụ của 30 hộ ở một khu dân cư trong một tháng như sau (*đơn vị: kW*):

50	47	30	65	63	70	38	34	48	53	33	39	32	40	50
55	50	61	37	37	43	35	65	60	31	33	41	45	55	59

Với các lớp: [30;35), [35; 40), …, [65;70].

7) Số cuộn phim mà 40 nhà nhiếp ảnh nghiệp dư sử dụng trong một tháng.

5	3	3	1	4	3	4	3	6	8	4	2	4	6
8	9	6	2	10	11	15	1	2	5	13	7	7	2
4	9	3	8	8	10	14	16	17	6	6	12		

Với các lớp: [0; 2], [3; 5], …, [15; 17].

8) Số người đến thư viện đọc sách buổi tối trong 30 ngày của tháng 9 ở một thư viện.

85	81	65	58	47	30	51	92	85	42	55	37	31	82	63
33	44	93	77	57	44	74	63	67	46	73	52	53	47	35

Với các lớp: [25; 34], [35; 44], …, [85; 94] (độ dài mỗi đoạn bằng 9).

9) Số tiền điện phải trả của 50 gia đình trong một tháng ở một khu phố (*đơn vị: nghìn đồng*)

Lớp	[375; 449]	[450; 524]	[525; 599]	[600; 674]	[675; 749]	[750; 825]
Tần số	6	15	10	6	9	4

10) Khối lượng của 30 củ khoai tây thu hoạch ở một nông trường (*đơn vị: gam*).

Lớp	[70; 80)	[80; 90)	[90; 100)	[100; 110)	[110; 120)
Tần số	3	6	12	6	3